



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Reviewing the review

**Citation for published version:**

Davis, CH, Bass, BL, Behrns, KE, Lillemoe, KD, Garden, OJ, Roh, MS, Lee, JE, Balch, CM & Aloia, TA 2018, 'Reviewing the review: a qualitative assessment of the peer review process in surgical journals', *Research integrity and peer review*, vol. 3, pp. 4. <https://doi.org/10.1186/s41073-018-0048-0>

**Digital Object Identifier (DOI):**

[10.1186/s41073-018-0048-0](https://doi.org/10.1186/s41073-018-0048-0)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Research integrity and peer review

**Publisher Rights Statement:**

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



RESEARCH

Open Access



# Reviewing the review: a qualitative assessment of the peer review process in surgical journals

Catherine H. Davis<sup>1,2</sup>, Barbara L. Bass<sup>2</sup>, Kevin E. Behrns<sup>3</sup>, Keith D. Lillemoe<sup>4</sup>, O. James Garden<sup>5</sup>, Mark S. Roh<sup>6</sup>, Jeffrey E. Lee<sup>1</sup>, Charles M. Balch<sup>1</sup> and Thomas A. Aloia<sup>1\*</sup>

## Abstract

**Background:** Despite rapid growth of the scientific literature, no consensus guidelines have emerged to define the optimal criteria for editors to grade submitted manuscripts. The purpose of this project was to assess the peer reviewer metrics currently used in the surgical literature to evaluate original manuscript submissions.

**Methods:** Manuscript grading forms for 14 of the highest circulation general surgery-related journals were evaluated for content, including the type and number of quantitative and qualitative questions asked of peer reviewers. Reviewer grading forms for the seven surgical journals with the higher impact factors were compared to the seven surgical journals with lower impact factors using Fisher's exact tests.

**Results:** Impact factors of the studied journals ranged from 1.73 to 8.57, with a median impact factor of 4.26 in the higher group and 2.81 in the lower group. The content of the grading forms was found to vary considerably. Relatively few journals asked reviewers to grade specific components of a manuscript. Higher impact factor journal manuscript grading forms more frequently addressed statistical analysis, ethical considerations, and conflict of interest. In contrast, lower impact factor journals more commonly requested reviewers to make qualitative assessments of novelty/originality, scientific validity, and scientific importance.

**Conclusion:** Substantial variation exists in the grading criteria used to evaluate original manuscripts submitted to the surgical literature for peer review, with differential emphasis placed on certain criteria correlated to journal impact factors.

**Keywords:** Surgery journals, Surgical research, Surgical manuscripts, Manuscript review, Journal reviewer

## Background

The first known incidence of peer review in the biomedical literature was performed by the Royal Society of Edinburgh in 1731, but the practice was not extensively adopted until the second half of the twentieth century [1]. Even then, the transition was far from organized and widely resisted by editors because of the low supply to demand ratio of medical articles [2]. Many medical journals first appointed small advisory boards and committees to perform "internal reviews" before transitioning to a more modern model of external peer review [2]. Over the course of the twentieth century, the number of medical articles submitted to

journals exceeded demand and has continued to exponentially grow. As of 2012, there were 2 million peer reviewed articles published annually by 28,000 scholarly journals [1]. In the current era, most of the scientific, including surgical, literature is evaluated and published via a peer review process.

While an optimal peer review process aims to maintain a high level of research integrity and a literature that supports the practice of evidence-based medicine, these systems are imperfect [1, 3]. Unfortunately, not all peer review is currently performed to a high standard, and there is concern that manuscripts published after inadequate peer review will negatively affect future literature reviews, meta-analyses, and most importantly, medical practice [1, 4]. Where there are additional hurdles to collection of high quality evidence in the form of randomized

\* Correspondence: [taaloia@mdanderson.org](mailto:taaloia@mdanderson.org)

<sup>1</sup>Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, 1400 Herman Pressler, Unit 1484, Houston, TX 77030, USA  
Full list of author information is available at the end of the article



© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

studies, less rigorous research could adversely affect surgical patient care [5].

Over the past 10 years, large strides have been made in creating clear and widely accepted guidelines for authors performing studies and writing manuscripts, specifically by formatting the reporting of various study designs [4–9]. Despite supposed widespread acceptance of such guidelines, adoption has been poor in the surgical literature [10]. In contrast, no consensus guidelines have been emerged to define the optimal criteria for editors to grade the peer reviewed literature. The only standardized guidelines for performing scientific peer review were published by the Council of Biology Editors (CBE, now the Council of Science Editors) in 1983 [11]. These guidelines emphasized grading specific manuscript components, including scientific importance, clarity, study design and methods, validity of interpretation and statistical methods, appropriate literature review, written presentation, and quality. The CBE recommendations were adapted and expanded by Frank in 1996, adding that reviewers should disclose conflicts of interest and editors should provide comprehensive instructions to reviewers, with an emphasis on training first-time reviewers [12]. Likewise, the only published guidelines for peer review that are specific to surgery [13–15] focus on educating student and novice peer reviewers on the process and largely mirror the CBE recommendations with two [13, 15] including more comprehensive instruction on proper reviewer report writing.

In the absence of clear consensus guidelines, editorial boards have evolved unique and variable styles for the grading of surgical manuscripts. To determine the degree of variability and any associations between grading components and journal impact factor, this project was designed to assess the current metrics used in the surgical literature to evaluate submitted original manuscripts.

## Methods

Manuscript grading forms for 14 of the highest circulation general surgery-related journals were collected from journal editors and peer reviewers (*Annals of Surgery*, *Annals of Surgical Oncology*, *BJS* (formerly the *British Journal of Surgery*), *Digestive Surgery*, *European Journal of Surgical Oncology*, *HPB*, *JAMA Surgery*, *Journal of the American College of Surgeons*, *Journal of Gastrointestinal Surgery*, *Journal of Surgical Oncology*, *Journal of Surgical Research*, *Surgery*, *Surgical Endoscopy*, and *World Journal of Surgery*). Journals were ranked in quality as determined by journal impact factor (2015, Thomson Reuters [16]). Each grading form was evaluated for content, including the type and number of specific quantitative and qualitative questions asked of the peer reviewer. Every question contained in the grading forms was recorded and categorized. This analysis included identification of specific questions related to the following: recommendation for publication/further steps,

overall manuscript rating, manuscript priority, need for statistical review, comments to author and editor, questions about specific manuscript components (abstract, background, methods, statistics, results, discussion), ethical issues, figures/tables, references, written presentation/grammar, conflict of interest, novelty/originality, clinical importance/relevance, scientific importance, and scientific validity. The inclusion of specific question types on reviewer grading forms was compared between the seven surgical journals with the higher impact factors and the seven surgical journals with lower impact factors using Fisher's exact tests. A two-tailed  $p < 0.05$  was considered statistically significant. All statistical analyses were performed using SPSS software (version 23.0, IBM Corp).

## Results

Overall, the median impact factor for the cohort of 14 journals included in the study was 3.25 (range 1.73–8.57). For the group of higher impact factor journals, the median impact factor was 4.26 (range 3.18–8.57), and for the group of lower impact factor journals, the median impact factor was 2.81 (range 1.73–3.15). Substantial variability in the distribution of the total number of questions was observed between journal groups. High impact factor group journals asked a median 11 questions (range 5–26) and lower impact factor journals asked a median of 10 questions (range 3–14). None of the 14 journals conducted blinded review (i.e., reviewers were aware of authorship at the time of their review). One of the 14 journals offered the reviewer the option of unblinding their name to the authors, but none mandated disclosure of the reviewer's name to the authors.

The content of the grading forms was also found to vary considerably. (Table 1). The only two questions asked by all 14 journals were overall recommendation and comments to the editor. All 7 of the higher impact journals solicited comments to the authors. Relatively few journals asked reviewers to grade-specific components of a manuscript, including the abstract, introduction, methods, results, discussion, figures and tables, references, ethical issues/IRB, and written presentation and grammar (range 2–5 out of all 14 journals). Interestingly, no lower impact journals included questions about the need for further statistical review, ethical issues, or reviewer conflict of interest.

Higher impact factor journal manuscript grading forms more frequently addressed these issues, including statistical analysis (5/7 vs. 0/7,  $p = 0.021$ ). Though not reaching statistical significance, possibly secondary to small group size, ethical considerations (3/7 vs. 0/7), and conflict of interest (2/7 vs. 0/7) trended to favor higher impact journals. In addition, higher impact factor journals tended to more commonly ask the review to assign a manuscript priority (4/7 vs. 2/7). In contrast, lower impact

**Table 1** Surgical manuscript review questions stratified by the journal impact factor

Question type	All journals	Higher impact journals	Lower impact journals	P value
Rating questions				
Recommendation	14/14	7/7	7/7	ND
Overall manuscript rating	4/14	2/7	2/7	ND
Manuscript priority	6/14	4/7	2/7	0.592
Statistical review required	5/14	5/7	0/7	0.021
Comments to author	13/14	7/7	6/7	1.000
Comments to editor	14/14	7/7	7/7	ND
Quantitative questions/specific manuscript components				
Abstract	2/14	1/7	1/7	ND
Introduction	3/14	1/7	2/7	1.000
Materials and methods	3/14	1/7	2/7	1.000
Ethical issues/IRB approval	3/14	3/7	0/7	0.192
Analysis/results	3/14	1/7	2/7	1.000
Discussion and conclusions	3/14	1/7	2/7	1.000
Figures/tables	4/14	2/7	2/7	ND
References	3/14	1/7	2/7	1.000
Written presentation/grammar	5/14	2/7	3/7	1.000
Conflict of interest	2/14	2/7	0/7	0.462
Qualitative questions				
Novelty/originality	7/14	2/7	5/7	0.286
Clinical importance/relevance	3/14	2/7	1/7	1.000
Scientific importance	7/14	3/7	4/7	1.000
Scientific validity	5/14	2/7	3/7	1.000

ND no difference

factor journals tended to more commonly request reviewers to make qualitative assessments of novelty/originality (5/7 vs. 2/7). Lower impact factor journals also asked more often about specific manuscript sections such as the abstract, introduction, methods, results, and discussion.

## Discussion

Although this study is limited by a small sample size, it did include widely circulated, English-language general surgery journals. The main findings of the study are that there is a significant variability in the specific criteria that surgical journals use to evaluate submitted original manuscripts and that the use of certain criteria is associated with impact factor levels. These data are congruent with previous literature on this topic, finding agreement on the wide variability in peer review structure [1, 4, 12, 17, 18].

Despite multiple critiques of the current system [3, 4, 12, 18, 19], implementation of consensus guidelines for standardization and optimization of the peer review process continues to be lacking. In this study, multiple of the specific points from the guidelines published by the CBE [11] and Frank [12] were identified in the reviewer

grading forms, but there was considerable inconsistency between different journals, and almost none addressed the majority of the suggested guidelines.

Regarding previous studies aimed at assessing the peer review process, two similarly designed but non-surgical and now dated studies have been previously reported in the scientific literature. First, Frank in 1996 studied the review forms, instructions for reviewers, and cover letters of 67 out of the top 100 journals rated by the 1989 Institute for Scientific Information citation frequency index [12]. As in the current study, Frank found great variability in the instructions and questions asked of reviewers. Only 25% of journals provided reviewer instructions longer than one page in length. Few journals asked for assessments of soundness and quality (29%) and ethical issues (36%) or reminded reviewers of confidentiality (46%). Half of the journals solicited assessments of manuscript conclusions (51%) or appropriateness of the manuscript for the particular journal (51%). Many journals asked for assessment of manuscript novelty (72%) and priority/importance/significance (88%). Almost all journals asked reviewers to recommend manuscript acceptance or rejection (96%). Second, Weller and colleagues in 1990 compared the

peer review process between 16 higher impact medical journals and 73 lower impact medical journals and found that lower impact journals were more likely to use peer reviewers in important decision points in the publication process than higher impact journals, which relied more on their editorial staff [17]. Lower impact journals were also more likely than higher impact journals to conduct blind review. Though this study provided an important snapshot of the peer review process, no reviewer feedback was collected as part of this study.

The issue of differences in the review process of journals with similar topic/focus areas (e.g., general surgery) has two sides. On the one hand, the data indicate that there may be room for improvement through standardization of the review process of the surgical literature. Although not providing concrete guidelines, previous authors have made suggestions for improvement of the peer review process, including formal training of reviewers by the journal [18, 19], standardization and creation of manuscript review protocols [12, 18], recruiting peer reviewers in the same niche field addressed in the manuscript [4], recognizing and treating peer review as a professional skill which is rewarded [3, 4, 18], editorial review of reviewer quality with feedback on performance to reviewers [18], and blinding of reviewers [18]. Given that the current study demonstrated none of the 14 top impact factor general surgery journals had a blinded review process, this could be an important advance in the surgical literature. There is mixed evidence in the literature supporting [20, 21] versus invalidating [22, 23] the importance of blinded peer review. However, it is possible that, as researchers in the same subspecialties have various professional relationships and in some cases may be competing for various grants or positions, non-blinded review may create significant bias [4].

On the other hand, variability in peer review grading criteria allows a journal and its editorial board to differentiate the focus and content of their publications, targeting the needs of their specific readership. Higher impact factor journals may justifiably place a greater emphasis on novelty and importance. Thus, while some basic criteria may be important for all journal review forms, there may be good reason to allow for diversity of emphasis or additional criteria. Along the same lines, rigid adherence to a relatively large number of criteria or check boxes, especially for an inexperienced reviewer, risks the reviewer missing the “forest for the trees.” Furthermore, editors may find the narrative detailed review in prose submitted by the reviewer to be more valuable than a quantitative scoring of individual manuscript components.

In order to reconcile these arguments that favor continued variability in the peer review grading process, it may be incumbent on editors to become more transparent to authors and readers about their journal's unique areas of emphasis (Table 2). Currently, the peer review process,

**Table 2** Suggestions for improvement of the manuscript review process in surgery

- 
- 1) Blinding reviewers to authors
  - 2) At least voluntary unblinding of reviewers to authors
  - 3) Continued emphasis on quantitative scoring of key manuscript elements including statistical validity and ethical concerns
  - 4) Transparency in unique criteria of emphasis to reviewers and authors
  - 5) Guidelines for structuring narrative comments to editors and authors [13]
- 

including the selection of reviewers, is incredibly variable between journals and is rarely transparent to the authors or the journal readership [1, 3, 12]. While different journals may necessarily emphasize different aspects in manuscript grading based on journal focus, process variability could also indicate that some journals are missing important specific elements of the review process that could improve the scientific quality of their publications [12].

The major concern regarding variability in the peer review process is the potential for this variability to introduce bias into the system. Peer review bias is an important problem in the scientific publication process. Peer review can be politically charged and bias, which can be positive or negative, can be directed for or against specific authors, fields, or institutions, resulting in publication bias and/or prolonged delays in publication time [1, 19]. Even as new markets open up within countries, fields, and the internet, as the academic credit toward promotion and tenure awarded for participation as an ad hoc reviewer to journals has declined, the available cohort of willing reviewers is shrinking [4]. This contraction of the reviewer pool is occurring even as the volume of submissions to surgical journals is increasing, particularly from international authorship groups [4]. Development of reviewer education programs and structured review forms that expedite the workload of reviews may encourage reviewers to stay engaged with the process and stave off the increasing pressure on editors to administratively reject submissions without peer review.

The data from this analysis suggest that overall variability and specific variability associated with impact factor may be limiting the overall quality in the surgical literature. While maintaining focus areas specific to a given journal, there may be an opportunity to increase uniformity and transparency of grading criteria through the acceptance of previously published guidelines. As there is no existing organization that brings together all of the editors of surgical journals, an organization such as the American College of Surgeons or the American Board of Surgery, which oversee all surgical specialties, may be well positioned to recommend standardized review practices to impact the quality of the surgical literature. Future specific areas for improvement of peer review of the surgical literature may



include standardization of the grading form, blinding of reviewers to authors, exposure of reviewer identity to authors, and increased transparency of the review process to submitting authors (Table 2).

## Conclusions

Substantial variation exists in the grading criteria used to evaluate original manuscripts submitted to the surgical literature for peer review, with differential emphasis placed on certain criteria correlated to journal impact factors. By directly comparing the review processes of similar journals, it is apparent that potential gaps exist in the review process and that the surgical literature may benefit from a more uniform template for literature peer review.

## Abbreviations

CBE: Council of Biology Editors

## Availability of data and materials

Available upon request to the corresponding author.

## Authors' contributions

CHD, JEL, CMB, and TAA contributed to the study design and concept of the manuscript. CHD and TAA collected and analyzed the data. All authors drafted the manuscript and were responsible for the critical edits of the manuscript. All authors read and approved the final manuscript.

## Ethics approval and consent to participate

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Author details

<sup>1</sup>Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, 1400 Herman Pressler, Unit 1484, Houston, TX 77030, USA.

<sup>2</sup>Department of Surgery, Houston Methodist Hospital, Houston, TX, USA.

<sup>3</sup>Department of Surgery, Saint Louis University School of Medicine, Saint Louis, MO, USA. <sup>4</sup>Department of Surgery, Massachusetts General Hospital, Boston, MA, USA. <sup>5</sup>Royal Infirmary of Edinburgh, Edinburgh, UK. <sup>6</sup>Orlando Health, The University of Florida Health Cancer Center, Orlando, FL, USA.

Received: 31 January 2018 Accepted: 9 May 2018

Published online: 23 May 2018

## References

- Manchikanti L, Kaye AD, Boswell MV, Hirsch JA. Medical journal peer review: process and bias. *Pain Physician*. 2015;18(1):E1–E14.
- Burnham JC. The evolution of editorial peer review. *JAMA*. 1990;263(10):1323–9.
- Patel J. Why training and specialization is needed for peer review: a case study of peer review for randomized controlled trials. *BMC Med*. 2014;12:128.
- Stahel PF, Moore EE. Peer review for biomedical publications: we can improve the system. *BMC Med*. 2014;12:179.
- McCulloch P, Feinberg J, Philippou Y, et al. Progress in clinical research in surgery and IDEAL. *Lancet*. 2018;
- Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol*. 2009;62(10):e1–34.

- Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c869.
- von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*. 2007;370(9596):1453–7.
- Larson EL, Cortazar M. Publication guidelines need widespread adoption. *J Clin Epidemiol*. 2012;65(3):239–46.
- Agha RA, Barai I, Rajmohan S, et al. Support for reporting guidelines in surgical journals needs improvement: a systematic review. *Int J Surg*. 2017;45:14–7.
- Committee CSM. CBE style manual. 5th ed. Bethesda: Council of Biology Editors, Inc.; 1983.
- Frank E. Editors' requests of peer reviewers: a study and a proposal. *Prev Med*. 1996;25(2):102–4.
- Stahel PF, Moore EE. How to review a surgical paper: a guide for junior referees. *BMC Med*. 2016;14:29.
- Seals DR, Tanaka H. Manuscript peer review: a helpful checklist for students and novice referees. *Adv Physiol Educ*. 2000;23(1):52–8.
- McFadden DW, LeMaire SA, Sarr MG, Behrns KE. How to review a paper: suggestions from the editors of surgery and the journal of surgical research. *Surgery*. 2017;162:1–6.
- 2015 Journal Citation Reports® Science Edition (Clarivate Analytics, 2017). <http://jcr.incites.thomsonreuters.com/JCRJournalHomeAction.action>.
- Weller AC. Editorial peer review in US medical journals. *JAMA*. 1990;263(10):1344–7.
- Smith R. Peer review: a flawed process at the heart of science and journals. *J R Soc Med*. 2006;99(4):178–82.
- Benos DJ, Bashari E, Chaves JM, et al. The ups and downs of peer review. *Adv Physiol Educ*. 2007;31(2):145–52.
- McNutt RA, Evans AT, Fletcher RH, Fletcher SW. The effects of blinding on the quality of peer review. A randomized trial. *JAMA*. 1990;263(10):1371–6.
- Godlee F, Gale CR, Martyn CN. Effect on the quality of peer review of blinding reviewers and asking them to sign their reports: a randomized controlled trial. *JAMA*. 1998;280(3):237–40.
- van Rooyen S, Godlee F, Evans S, Smith R, Black N. Effect of blinding and unmasking on the quality of peer review: a randomized trial. *JAMA*. 1998;280(3):234–7.
- Chung KC, Shauver MJ, Malay S, Zhong L, Weinstein A, Rohrich RJ. Is double-blinded peer review necessary? The effect of blinding on review quality. *Plast Reconstr Surg*. 2015;136(6):1369–77.

### Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

